EPFL

Unil
UNIL | Université de Lausanne

Digital traces, navigational paths, and intellectual mobilities
How researchers navigate libraries and online platforms and how to foster such practices

Traces numériques, parcours de navigation et mobilités intellectuelles
Comment les chercheurs naviguent dans les bibliothèques et les plateformes en ligne

26-27 November 2020

Please contact simon.dumasprimbault@epfl.ch to obtain the Zoom meeting room id.
Veuillez contacter simon.dumasprimbault@epfl.ch afin d'obtenir l'identifiant de réunion Zoom.

Thursday, November 26, *afternoon* | Jeudi 26 novembre, *après-midi*

14h00 – 14h10	Introductory speech | Introduction Simon Dumas Primbault, Jérôme Baudry, Jean-François Bert
14h10 – 15h00	Marc Jahjah (Université de Nantes) « Du surlignement au remembrage : gestes livresques, gestes textiels »
15h00 – 15h50	Irène Bastard (BnF) and Valérie Beaudouin (Télécom Paris) "Mixed methods in sociology: how to learn about online and offline uses of BnF?"
15h50 – 16h10	Break | Pause
16h10 – 17h00	Maud Ehrmann (EPFL) and Matteo Romanello (Unil) "Interactive exploration of the digital newspaper space with *impresso*"

Friday, November 27, *morning* | Vendredi 27 novembre, *matin*

9h00 – 9h50	Robert West (EPFL) "User navigation on Wikipedia"
9h50 – 10h40	Nicolas Baya-Laffite, Ogier Maitre, and Boris Beaude (Unil) "Discovering the world through collective attentions: navigating Wikipedia's archive of consultation traces"
10h40 – 11h00	Break | Pause
11h00 – 11h50	Christian Jacob (EHESS) et Axel Le Roy (Sorbonne Université) « *Savoirs*, de l'encodage aux algorithmes : générer des parcours de lecture dans une bibliothèque d'histoire et d'anthropologie des sciences et des techniques »
11h50 – 12h40	Giovanni Colavizza and Karlijn Roex (University of Amsterdam) "Netherlands-based science in times of COVID-19: First results from our survey"

Abstracts | Résumés

« Du surlignement au remembrage : gestes livresques, gestes textiels »
Marc Jahjah

Récemment lancé, le logiciel « LiquidText » a aussitôt été présenté comme « révolutionnaire » par la rhétorique traditionnelle du web contemporain. À l'inverse de tous ses concurrents conçus depuis les années 90, il permet certes d'annoter un texte, de le surligner, d'intégrer des commentaires mais, bien plus, de le découper, d'en déplacer les unités ainsi fragmentées dans un espace dédié, où il peut être de nouveau recomposé, recontextualisé, réélaboré, tout en étant techniquement et sémantiquement lié au texte d'origine. En cela, il s'inscrit dans une histoire longue des pratiques intellectuelles qui, dès le XIIᵉ siècle, ont fait du texte un silo, une mine dans laquelle puiser pour extraire des informations utiles (Ivan Illich, *Du lisible au visible*, 1991, Cerf). Dans cette communication, j'aimerais situer les logiciels informatiques destinés aux travailleurs du savoir, dans la « révolution scribale » dont ils sont des avatars actualisés et contemporains, à travers un arc conceptuel : fragmentation, instrumentalisation, anthologisation, industrialisation et spectacularisation du texte. Nous serons alors à même de comprendre que le monde scolastique et ses pratiques, loin de s'opposer à l'âge des systèmes et à l'extraction des ressources, en sont l'une des formes ajustées, qui passe par de menus gestes, des signes minuscules auxquels nous participons quotidiennement et dont nous faisons fonctionner l'impensé corporel, économique, social, idéologique et technique.

Marc Jahjah est maître de conférence à l'Université de Nantes. Ses travaux portent sur les pratiques savantes, documentaires, numériques, littéraires et créatives.

"Mixed methods in sociology: how to learn about online and offline uses of BnF?"
Irène Bastard, Valérie Beaudouin

The National Library of France (BnF) develops and offers to its users a set of services for online access and for the reading rooms based on its large and historical resources (legal deposit of printed productions or musics and multimedia, electronic resources, digitized collections, web archives, etc.). With 76.000 readers registered at François-Mitterrand or Richelieu libraries (reading rooms) and 16 millions of visits on Gallica (the digital library), the BnF is facing a major challenge: how to develop a strong knowledge of library users and uses in order to adapt the exploration tools and to contribute to the scientific and academic community? Do archives and manuscripts have the same signification for academic or amateur historians? How does digital access transform the practices of these communities? Furthermore, what kind of methodologies can we rely on to better understand users and their uses; which patterns should one follow and monitor to improve service?

Three research projects based on different methodologies will be mobilized to answer these questions: a log analysis of Gallica, a video-ethnography on Gallica users, and qualitative interviews with researchers on political sciences.

Irène Bastard holds a degree in digital technologies (Télécom Paris) and a PhD in Sociology. She conducts research on users experience for the BnF (National Library of France) Research and Stragegy department. Each year, they produce statistics on users' registrations, surveys about exhibitions and services, foresight on developments. As a sociologist working on digital activities, her recent studies deal with online and offline imbrications in researchers' practices.

Valérie Beaudouin is a Professor of sociology at Télécom Paris and an associate researcher at EHESS-CEMS (Centre d'étude des mouvements sociaux). She graduated from ENSAE, as a Statistician Economist and holds a PhD in Linguistics (EHESS) and a habilitation in sociology. She works on the impact of digitization on social practices. She specialized on digital methods for social sciences (text mining, network analysis, netnography). She has been collaborating with BnF on the study of library uses since 2013.

"Interactive exploration of the digital newspaper space with *impresso*"
Maud Ehrmann, Matteo Romanello

impresso. Media Monitoring of the Past[1] is an interdisciplinary research project in which a team of computational linguists, designers and historians collaborate on the datafication of a multilingual corpus of digitized historical newspapers. The primary goals of the project are to improve text mining tools for historical text, to enrich historical newspapers with (semi-) automatically generated data and to integrate such data into historical research workflows by means of a newly developed user interface[2].

In this presentation, we will focus on how the impresso app was designed to integrate into research workflows newspapers metadata and the data produced by various semantic enrichments (topic modelling, named entity processing, text reuse detection). Built around the idea of *interactively exploring* a large-scale newspaper corpus, the interface allows the users to move within the digital newspaper space in multiple directions, e.g., by looking at the data from different angles, by shifting scales (distant vs. close), and by inspecting materials through data-driven comparisons. While each interface component allows for a different "itinerary" through the data, their tight integration into a coherent design allows users to combine exploration paths into research workflows. We will introduce the impresso project, present the interface and illustrate its functionalities by means of real-world examples.

---

[1] https://impresso-project.ch/
[2] https://impresso-project.ch/app/#/

Maud Ehrmann is a research scientist at EPFL's Digital Humanities Laboratory. She holds a PhD in Computational Linguistics from the Paris Diderot University (Paris 7) and has been engaged in a large number of scientific projects centred on information extraction and text analysis, both for present-time and historical documents. Her main research interests span natural language processing and digital humanities, and include information extraction, historical document processing and text mining, named entity processing, multilingual and historical resources creation, NLP system evaluation, knowledge representation, and large-scale infrastructure. Her current work at the DHLAB focuses on 'impresso - Media Monitoring of the Past', a SNSF Sinergia project.

Matteo Romanello is Lecturer at the University of Lausanne, where he conducts a project on the commentary tradition of Sophocles' *Ajax*. Matteo is a Classicist and a Digital Humanities specialist with expertise in various areas of the Humanities, including archaeology and history in addition to classics. His main research interests include natural language processing, information extraction, citation mining/analysis, and applications of semantic web technologies to data in the humanities. After obtaining his PhD from King's College London, he worked as a research scientist at EPFL's DHLAB on the Linked Books and Impresso projects, before moving to his current position. He was also teaching fellow at the University of Rostock, researcher at the German Archaeological Institute, and visiting research scholar at Tufts University.

"User Navigation on Wikipedia"
Robert West

The World Wide Web is an intricate network of information resources. In order to derive value from the Web, users need to navigate it. Even in the presence of powerful search engines that can "teleport" users to distant destinations, page-to-page navigation remains crucial. At the EPFL Data Science Lab, we have been intensively researching how users navigate the Web, with a particular focus on Wikipedia, the world's largest encyclopedia. In this talk, I will give an overview of our work on user navigation on Wikipedia, describing insights from (1) navigational traces collected via the navigation game Wikispeedia.net, (2) Wikipedia's tera-byte-scale server logs, (3) client-side instrumentation that we have built in order to capture the traffic that Wikipedia drives to third-party websites, as well as (4) surveys where tens of thousands of respondents have reported their reasons for using Wikipedia.

Robert West is a tenure-track assistant professor of computer science at EPFL (the Swiss Federal Institute of Technology, Lausanne), where he heads the Data Science Lab. He received his PhD in Computer Science from Stanford University, his MSc from McGill University, Canada, and his undergraduate degree from Technische Universität München, Germany. His research aims to understand, predict, and enhance human behavior in social and information networks by developing techniques in computational social science, social network analysis, machine learning, and natural language processing. Bob also collaborates closely with the Wikimedia Foundation, in

his role as a Wikimedia Research Fellow. He is a co-founder of the Wiki Workshop and the Applied Machine Learning Days.

"Discovering the world through collective attentions: navigating Wikipedia's archive of consultation traces"

Nicolas Baya-Laffite, Ogier Maitre, Boris Beaude

Over its 20 years, Wikipedia has become a core piece of our digital, everyday-life information and knowledge ecosystem. When users turn to it, the matters that drew their attention are captured in Wikipedia's article consultation logs, which grasp like no other social media platform what arises interest at a given moment. From a scholarly perspective, Wikipedia's pageviews data dumps offer a massive, detailed and up-to-date archive of collective attention and memory dynamics. While open, common, and with no privacy constrains bearing upon their usage, these data remain barely exploited in the humanities and the social sciences. This is astonishing considering their impressive potential for better understanding fundamental aspects of collective life across time, space and cultures. A reason for this has to do with such archive not being directly amenable for non-aprioristic exploration – something that limits the discovery potential of navigating the archive without a specific question in mind. So, how can we make consultation logs navigable in such a way that the scientific imagination be encouraged through serendipity? This is the challenge that the WikiMaps interdisciplinary project has tackled by developing an innovative navigation platform of consultation logs. Repurposing a 7-terabyte database comprising one thousand billion consultation log lines for the 300 linguistic editions, the WikiMaps current approach proposes to focus on consultations of pages identified with geocodes. Through a cartographic projection, specific views of the world and the ways in which we pay attention to places become freely explorable in a multiscale perspective from the planisphere to the street. Differential geographies emerge from the unequal interest elicited by content referring to what is taking place in the world over time. To contribute to the workshop's theme, we will present the navigational modes of the device and reflect upon the conceptual and technical aspects of its development, its heuristic value, and its use. This last point is then the occasion for imagining a further reflexive step: namely, documenting the navigation paths within the WikiMaps' device in order to produce a topology of mobility through such a world of collective attentions.

Nicolas Baya-Laffite is a researcher and lecturer at the STSlab at the Université de Lausanne (UNIL). He holds a PhD in Science and Technology studies from the EHESS, Paris. His research deals with digital transformations, their impact on knowledge production, and controversies thereon. Within the WikiMaps project, his work focuses on the repurposing of Wikipedia's digital traces. Previously, he worked on controversy mapping at the médialab Sciences Po.

Ogier Maitre is a researcher at the STSLab of the UNIL. He holds a PhD in computer science from EPFL. His research deals with data mapping and visualization, social systems modeling and

representation methods. Within the WikiMaps project, he worked on front-end and back-end development. Previously, he worked at the Chôros Lab of the EPFL where he collaborated with Boris Beaude in the development of a prototype of the WikiMaps navigation device.

Boris Beaude is an assistant professor in digital humanities, culture and society at the STSlab, at the Univeristy of Lausanne. He is the lead of the WikiMaps project. His research deals with the Internet as a singular space of coexistence. His recent work focuses on the paradigm shifts that occur with the renewal of the conditions of access to knowledge of individual and collective practices. Before joining UNIL, he was a researcher at the Chôros Lab of the EPFL.

« *Savoirs*, de l'encodage aux algorithmes : générer des parcours de lecture dans une bibliothèque d'histoire et d'anthropologie des sciences et des techniques »
Christian Jacob, Axel Le Roy

Savoirs est une bibliothèque intelligente, dynamique et numérique en cours de développement consacrée à l'histoire et à l'anthropologie des sciences et des savoirs. Construite dans une perspective comparatiste et interdisciplinaire, elle explore l'idée que le numérique permet le rapprochement multidimensionnel des textes et des idées. Comme Aby Warburg l'avait pressenti, théorisé et mis en pratique au seuil du vingtième siècle une bibliothèque est un dispositif heuristique qui aide à la pensée, à l'imagination, à la découverte. S'il est des bibliothèques que l'on visite pour trouver ce que l'on connait déjà, l'agencement physique des livres dans les étagères a le pouvoir d'influencer la construction des savoirs. Mais aujourd'hui, que nous permettent les technologies numériques pour le référencement des textes ? pouvons-nous espérer lire autrement ? comment répondre aux attentes des lecteurs et créer de la digression, de la surprise ? En revenant sur l'histoire du projet nous présenterons les acteurs, les pratiques et méthodes mobilisées dans la bibliothèque Savoirs pour mettre les textes en dialogue et répondre aux attentes des lecteurs sans négliger la fabrication d'une sérendipité numérique, permettant d'aller au-delà des stratégies de suggestion du même, familières aux internautes contemporains.

Christian Jacob est directeur de recherche au CNRS et directeur d'études à l'EHESS. Historien des savoirs, dans l'Antiquité classique et dans une perspective comparatiste. Porteur du projet *Savoirs*.

Axel Le Roy est doctorant contractuel à la Faculté des lettres de Sorbonne Université (LabEx Obvil) et chercheur associé à la Bibliothèque nationale de France (département des Sciences et techniques). Sa thèse porte sur la construction des savoirs sur les animaux chinois dans l'Europe du XVIIIe siècle.

"Netherlands-based science in times of COVID-19: First results from our survey"
Giovanni Colavizza, Karlijn Roex

The SARS-CoV-2 pandemic generated an 'infodemic' (WHO, 2020): a sea of information that contains a hard-to-distinguish portion of rumours and misinformation. Scientists play a vital role in generating and distributing reliable information. Between June and August 2020, we collected survey responses from scientists based at Dutch research institutions. We used a multi-method stratified sampling strategy. We gathered 777 responses, of which 552 (71%) completed the survey at least up to 75%. This sample is strongly representative in terms of academic position (e.g., PhD student, post doc, assistant professor, etc.) and age. Women are over-represented (50% of our sample, while only 40% of academic researchers). We found that a quarter of our respondents (24%) is involved in research that is related to the COVID-19 pandemic. A small subset (16%) had already been studying pandemics or viruses before. These were mostly working in public health, the medical sciences or biological sciences. 'New' COVID-19-researchers (N = 120) were often working in public health and the medical sciences (38%) or the social sciences (26%). Finally, a quarter (24%) of all researchers that had encountered COVID-19 misinformation at online fora, had responded to this by trying to correct it. However, we found a gender imbalance in the propensity to address misinformation online. This corresponds with previous findings on gender inequalities in scholarly publishing and media coverage of experts on COVID-19. Note that these findings are descriptive and not yet inferential.

Giovanni Colavizza is Assistant Professor of Digital Humanities at UvA, visiting researcher at The Alan Turing Institute and at the Centre for Science and Technology Studies (CWTS), Leiden University. He did his PhD at the Digital Humanities Laboratory of the EPFL in Lausanne, working on methods for text mining and citation analysis of scholarly publications. Giovanni is interested in several topics spanning from AI for cultural heritage (part of UvA CREATE), to crypto art markets and the public understanding of science.

Karlijn Roex is a postdoctoral researcher at the University of Amsterdam. Her research focuses on science communication and 'infodemics' during times of crisis - such as the present COVID-19 crisis (together with Giovanni Colavizza). In 2018, she received her PhD on suicide rates during economic crises at the University of Cologne and the Max Planck Institute.